October 28, 1964

Dr. S. H. Eisman
United States Army
Commanding Officer
Frankford Arsenal
Philadelphia, Pa.

Dear Dr. Eisman:

I was very pleased that you got in touch with me. Since I came across your paper in the Journal of Chemical Documentation, and Gene Garfield had also called this to my attention, I had been meaning to get in touch with you but thought I should put some final touches on the DENDRAL system, and this stretched out without my realizing it.

My principle concern has been the derivation of a canonical form for any chemical graph and to do this as far as possible on the basis of elementary topological principles. Blair and Henze, in a bit of work that seems to have been almost totally overlooked (I certainly wish I had a Citation Index for the intervening years), showed us one pretty good way to do this for tree structures and linear DENDRAL offered no very fundamental problems. The main issues are (1) to find the centroid of the tree, and (2) to evaluate each of the twigs from each subsequently branched node to apply a canon of preference. Most of our programming has been done on the IBM 7090 with Balgol (in fact the recursive facilities of Algol on our Burroughs 5000 also prove to be extremely useful). It quickly turned out that the evaluation of a twig bears a close analogy to the processing of a stack such as is done during compilation for the translation of arithmetic expressions. So some very handy chunks could be extracted bodily from the compiler programs for this purpose. We worked out one format of representation of the trees that seem to us reasonably convenient, but this is relatively incidential to my principle concern and I would leave it to experience to tell just which coding system best meets a variety of needs. I am still in

EISMAN

the course of polishing up a description of the details of the analysis, but what I have just summarized should give you a reasonable picture of it and I include a sheet or two with some examples.

Most of our troubles, of course, have been with complex cyclic structures and it took quite a while before we hit on the most obvious and most elementary way of dealing with them. I think the enclosed manuscript should be a sufficient explanation. Let me add that each fundamental figure has a standard dequence of paths so that a statement of which fundamental figure is used for the mapping and then a list of the paths describes the structure. For a canonical mapping one must, of course, pay close attention to the symmetries of the figure and most of the complications arise from this source once the scheme is understood. I will send you the more detailed exposition as soon as I have been able to put it into reasonably literate English.

The declaration and definition of a ring gives meaning to a label which can then be used as a node in linear DENDRAL. In the precedence relationships the order of a ring counts first. Thus, in general, the centroid of a partly connected graph will be one of the rings. The only exception to this would arise when two equal subgraphs are joined by a linear path and such a graph is then treated like a symmetrical tree. However, many of these details really are quite optional and I feel that it would be quite important to hammer out what the best way of handling them really is. I have made one definite set of specifications just to have some point of departure and to lend some air of concreteness to the general proposal.

My principle aim in this work was not so much to generate a notational system as to provide a framework whereby a computer program could generate "all possible organic molecular structures" seriatim without redundancy. It was only afterwards that I realized that this aim should also be a specification of any formal notational system.

I would, of course, be delighted if any use can be made of these ideas for further development. It was the certainty that they would soon be superseded that led me to call the present system "DENDRAL-64".

The complete implementation of such a system is far beyond both my capacity and my principle interests, and as I wrote before, I would like to use it mainly as a tool for processing chemicals deductions on the computer. The mathematical analysis, so far, is not very sophisticated - for example, I am really not altogether happy about the way that four vertices are taken care of and I think that Polya's approach to handling symmetries would be very powerful if I could learn how to use it. I have managed to stir only a very small interest in these questions among the local mathematicians but one Ph.D. student probably will spend some time in further studies of canonical representations of graphs. If there is anything I can clear up here from having given only such a sketchy presentation, please let me know. As Gene may have told you, I would be especially delighted if he would take on the challenge of integrating DENDRAL into his name translation program.

Sincerely yours,

Joshua Lederberg
Professor of Genetics

JL:as

Encl.